Selective Verification Strategy for Learning from Crowds (Supplementary Materials)

Tian Tian, Yichi Zhou, Jun Zhu
Tsinghua University

A Proof of Proposition 1

Proposition 1. *Define* $I_{\mathcal{L}}(\boldsymbol{W}) := \mathbb{E}_{\boldsymbol{x}\sim\mathcal{L}}[I(\boldsymbol{x}, \boldsymbol{W})]$ *and* $I_{\mathcal{X}}(\boldsymbol{W}) := \mathbb{E}_{\boldsymbol{x}\sim\mathcal{X}}[I(\boldsymbol{x}, \boldsymbol{W})]$*, the expected loss error of the semi-supervised estimate* $\mathbb{E}[G(\hat{\boldsymbol{W}}^{\mathcal{L}}, \mathcal{X}) - G(\boldsymbol{W}^*, \mathcal{X})]$ *with respect to* B *ground truths sampled from* $\mathcal{L}$ *is upper bounded by* $\mathcal{O}\Big(\big(1 + \frac{\lambda B}{N}\big) \operatorname{tr}\Big(\big(I_{\mathcal{X}}(\boldsymbol{W}^*) + \frac{\lambda B}{N} I_{\mathcal{L}}(\boldsymbol{W}^*)\big)^{-1} I_{\mathcal{X}}(\boldsymbol{W}^*)\Big)\Big)$.

Proof. We follow the notations of Chaudhuri et al. [1]. We denote

$$\psi_i(\boldsymbol{W}) := U(\boldsymbol{W}^*) - \lambda B \log p(y_i|\boldsymbol{x}_i, \boldsymbol{W}),\ \forall \boldsymbol{x}_i \in \mathcal{L},$$

where y_i is sampled from $p(y_i|\boldsymbol{x}_i, \boldsymbol{W}^*)$. When $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\top]$ exists and is positive definite, $\psi_i(\boldsymbol{W})$ is smooth and strong convex. We denote $P(\boldsymbol{W}) := \mathbb{E}[\psi_i(\boldsymbol{W})]$ and $Q(\boldsymbol{W}) := G_{\boldsymbol{X}}(\boldsymbol{W})$, and the latter is the expected loss when the distribution of the ground truths for all tasks are observed. We also have $\nabla Q(\boldsymbol{W}^*) = \boldsymbol{0}$.

The Hessian of the loss on one verification sample $\boldsymbol{x}$ is

$$\begin{aligned}\frac{\partial^2 G(\boldsymbol{W}, \boldsymbol{x})}{\partial \boldsymbol{W}^2} &= -N \cdot \frac{\partial^2 U(\boldsymbol{W})}{\partial \boldsymbol{W}^2} - \lambda B \cdot \frac{\partial^2 \log p(y|\boldsymbol{x}, \boldsymbol{W})}{\partial \boldsymbol{W}^2} \\ &= N \cdot \mathbb{E}_{\boldsymbol{x}\sim\mathcal{X}}[I(\boldsymbol{x}, \boldsymbol{W})] + \lambda B \cdot I(\boldsymbol{x}, \boldsymbol{W}) \\ &= N \cdot I_{\mathcal{X}}(\boldsymbol{W}) + \lambda B \cdot I(\boldsymbol{x}, \boldsymbol{W}).\end{aligned}$$

Then we directly apply the Lemma 1 of Chaudhuri et al. [1] on $G_{\mathcal{X}}(\boldsymbol{W})$, we have that

$$\mathbb{E}[G(\hat{\boldsymbol{W}}^{\mathcal{L}}, \mathcal{X}) - G(\boldsymbol{W}^*, \mathcal{X})] = \mathcal{O}\left(\left(1 + \frac{\lambda B}{N}\right) \operatorname{tr}\left(\left(I_{\mathcal{X}}(\boldsymbol{W}^*) + \frac{\lambda B}{N} I_{\mathcal{L}}(\boldsymbol{W}^*)\right)^{-1} I_{\mathcal{X}}(\boldsymbol{W}^*)\right)\right).$$

Here we ignore all the constants and small quantities, since we only care about the relationship between the expected loss error and the verification subset. □

Reference

[1] Kamalika Chaudhuri, Sham M Kakade, Praneeth Netrapalli, and Sujay Sanghavi. Convergence rates of active learning for maximum likelihood estimation. In *NIPS*, 2015.